



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

9940 572 462

6381 907 438

ijircce@gmail.com

www.ijircce.com

Network Anomaly Detection Based on Traffic Classification

Livya Susan Mathew, Meghna Mary Biju, Sinchu Biju, Sandhra Sanjeev V S, Parvathy M N

U.G. Student, Department of Computer Engineering, Mount Zion College of Engineering, Kadamanitta, Kerala, India

U.G. Student, Department of Computer Engineering, Mount Zion College of Engineering, Kadamanitta, Kerala, India

U.G. Student, Department of Computer Engineering, Mount Zion College of Engineering, Kadamanitta, Kerala, India

U.G. Student, Department of Computer Engineering, Mount Zion College of Engineering, Kadamanitta, Kerala, India

Associate Professor, Department of Computer Engineering, Mount Zion College of Engineering, Kadamanitta, Kerala, India

ABSTRACT: Data centers in higher education institutions, as well as those of large corporations, face challenges in terms of traffic flow management. In some cases, due to the limited hardware resources used for this purpose, and in others, despite having enough high-performance equipment, the centers lag behind when the traffic flow grows exponentially due to the memory limitations of the devices, which slows down the network performance. The contribution of this investigation work is the implementation of a classifying elephant and mice system using machine learning techniques for the early detection with the first flow based on the dynamic calculation of the threshold, according to the input parameters of the final system. In the first instance, training algorithms are used to determine the best performance, then the proposed algorithm determines the model with the best prediction, obtained from the supervised learning algorithm trained in off line mode. Finally in the phase of online prediction, the algorithm is capable of predicting with high precision the type of traffic in terms of the input flow, and updates in a dynamic way the threshold to determine whether the traffic is elephant or mice. With this information the network hardware can decide then to route the flows according to their characterization. According to the results, the model that best generates predictions is the decision tree with a 100% confidence level.

KEYWORDS: Network Intrusion Detection System, Data pre-processing, real-time monitoring.

I. INTRODUCTION

The number of dangers to users' information has increased as the Internet has expanded. The threats vary from minor annoyances to life-threatening situations. Furthermore, in recent years, the assault tools and methodologies have become significantly more complex. The Network Intrusion Detection System (NIDS), as the protection mechanism behind the firewall, must reliably identify malicious network attacks, provide realtime monitoring and dynamic protective measures, and make strategic decisions. The intrusion detection under huge pressure due to hiding of malicious or infected data hiding among large number of normal data and even machine learning algorithm comes under constrains and makes mistakes by allowing infected data through defence line to PC. Since Lecun et al introduced the theory of Deep Learning as an important and vital subfield of machine learning, deep learning has shown excellent performance and played big role. Deep Learning is a subfield of artificial intelligence that is built on the discipline of machine learning. Deep learning will suffice since neural networks imitate human brain. Nothing is explicitly programmed in deep learning. Essentially, it is a machine learning class that does feature extraction and transformation using a significant number of nonlinear processing units. Each successive layer accepts the results from the previous layer as input. The main objective of network management is to preserve network availability and improve performance. Network management is becoming a challenge with the growth in network size, traffic volume and the diversity of Quality of Service (QoS) requirements. There is a wide range of applications with different requirements and constraints on network resources. Another aspect to highlight is the routing of packets, given the same network planning. A major problem is the early classification of traffic (mice or elephant), so the network hardware can make decisions when routing flows depending on the nature of the traffic itself. Based on the latter, it is expected that once the packets arrive at the switching systems, they will be able to identify what type of traffic it is with the first arrival flow and decide in a more optimal way how to route the traffic more efficiently.

II. RELATED WORK

There are a few different ways to avert digital attacks, one of which is by utilizing Intrusion Detection Systems. IDS is one segment of system security that ensures information and data security, by checking the traffic on a bundle of information to identify an interruption or anomaly. Finally, the deep local feature information extracted from the improved convolutional layer is vectorized into the form of a capsule, which can more accurately model the details of road network attributes and features and improve the model expression power and prediction performance. The network is tested on the Wenyi Road dataset and the public 23 dataset SZ-taxi. Compared with other models, the evaluation indicators of MCapsNet are better than other models in the tests of different time periods and predictors. The IDS used in this study is anomaly-based. The anomaly detection or outlier detection system assume that abnormal behavior is malicious. The idea is to train the machine learning model to learn normal behavior and then look for abnormal behavior or anomalies and raise alerts accordingly.

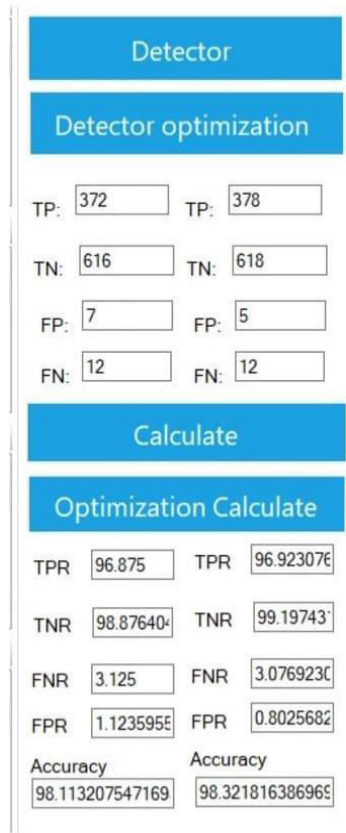
III. METHODOLOGY

Real-world data frequently contains noise, missing values, and is in an unsatisfactory format that cannot be used directly in machine learning models. Data pre-processing is performed first in our intrusion detection structure in the proposed intrusion detection model. Data pre-processing is a major prerequisite for cleaning data and making it fit for a machine learning model, which enhances the model's effectiveness and precision. We remove null rows and columns and remove duplicative values. Then we separate data into training set (subset to train the model) and testing set (a subset to test the train model). The majority of the data is used for training, while a smaller fraction is used for testing. There are 41 features in the Dataset. 30 features contain continuous values. The data set contains integers as well as floating-point numbers. Several char values require some preprocessing. These were assigned binary values through one-hot encoding and thus the non-numeric values were converted to numeric values, which can be given as the input to the algorithms. Then analyzing the average depth required to isolate each point.

1)DOS: This stands for Denial of service attack. It causes the server to become unavailable by bombarding it with false requests. These attacks are very prevented in IoT and wireless networks. It can occur in the transport layer as well as the application layer. 2)Probe: In this attack, the attacker sweeps through various hosts and services to identify open ports. 3)U2R: User To Root attacks are less common as compared to Dos attacks. The attacker gains access and attempts to gain root privilege. 4)R2I: It stands for Remote to local attack. In this attack, generally known to be propelled by an attacker to increase unapproved and remote access to a victim client machine in the whole system.

IV. EXPERIMENTAL ANALYSIS

After training the machine learning models, it is observed that 98.32% corresponds to mice traffic and 5.11% to elephant traffic. Once the independent and dependent variables were defined, the data were preprocessed and normalized. For this purpose, the K-fold cross-validation procedure was used, which is a standard method for estimating the performance of a machine learning algorithm on a set of data. The best-performing model-based confusion matrix was the decision tree classification algorithm, which scored 100% for predictions. The algorithm can predict the type of traffic depending on its size. Similarly, if the size is smaller, it will be identified as mice traffic. Data security wouldn't be complete without a solution to backup your critical information. Though it may appear secure while confined away in a machine, there is always a chance that your data can be compromised. You could suddenly be hit with a malware infection where a virus destroys all of your files. Someone could enter your computer and thief data by sliding through a security hole in the operating system. Perhaps it was an inside job that caused your business to lose those sensitive reports. If all else fails, a reliable backup solution will allow you to restore your data instead of starting completely from scratch. It is important to note that, by default, most of the traffic in the network is usually mice traffic. In addition, the algorithm shows high sensitivity, around 100%, allowing it to discriminate the negatives and positives of the trained instances correctly when predicting.



Detector

Detector optimization

TP: TP:

TN: TN:

FP: FP:

FN: FN:

Calculate

Optimization Calculate

TPR TPR

TNR TNR

FNR FNR

FPR FPR

Accuracy Accuracy

(a)

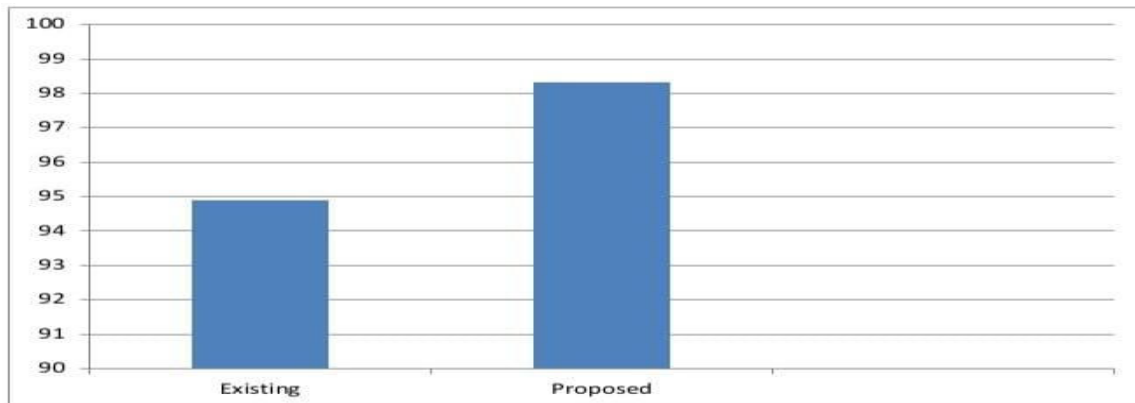
Fig.(a): Optimization calculation of existing system and proposed system

The Experimental results of based on the dataset

Model Name	Accuracy
Existing System	94.89
Proposed System	98.32

(b)

Fig.(b). Experimental results based on the dataset



(c)

Fig.(c): Accuracy Graph

V. CONCLUSION

Anomaly detection has great promise in this area, as it is efficient to train and detects anomalies with low false positive and false negative rates. In the implementation, it has been found that the anomaly detection process can be improved using various values of the available parameters for these algorithms. Also, it could be concluded that a more complete and clean data set leads to better results. The encryption purpose system uses a password-based AES algorithm which contains iteration, salt & provides MD5 hashing on encrypted IP address. The existing system & very few authors also focused on user sensitive information but this system is more strong because it provides encryption & hashing on sensitive information & also detects insider DDoS attack. In the proposed work, a secure log system is designed which will provide secure and reliable logs to investigators for cloud forensics. The Secretness and Privacy of cloud users will be preserved by using a searchable encryption technique. Modification of logs by server is not possible as logs are fully encrypted. This work can be extended to detect different cloud attacks like man-in-the-cloud attack, insider attack or other available log analyzer programs for detecting cloud attacks and helping in cloud forensics procedures.

REFERENCES

- [1] M. Afaq, S. U. Rehman, and W.-C. Song, "Visualization of elephant flows and QoS provisioning in SDN-based networks," in Proc. 17th Asia-Pacific Netw. Oper. Manag. Symp. Manag. (APNOMS), Aug. 2015, pp. 444–447.
- [2] P. Jurkiewicz, "Boundaries of flow table usage reduction algorithms based on elephant flow detection," in Proc. IFIP Netw. Conf. (IFIP Netw.), Jun. 2021, pp. 1–9, doi: 10.23919/IFIPNetworking52078.2021.9472832.
- [3] S. Floyd, "Simulation is crucial," IEEE Spectr., vol. 38, no. 1, p. 76, Jan. 2001.
- [4] C.-Y. Lin, C. Chen, J.-W. Chang, and Y. H. Chu, "Elephant flow detection in datacenters using OpenFlow-based hierarchical statistics pulling," in Proc. IEEE Global Commun. Conf., Dec. 2014, pp. 2264–2269.
- [5] M. E. Crovella, "Performance evaluation with heavy tailed distributions," in Proc. Int. Conf. Modelling Techn. Tools Comput. Perform. Eval., vol. 2221, 2001, pp. 1–10.
- [6] C. Estan and G. Varghese, "New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice," ACM Trans. Comput. Syst., vol. 21, no. 3, pp. 270–313, Aug. 2003.
- [7] K.-C. Lan and J. Heidemann, "A measurement study of correlations of internet flow characteristics," Comput. Netw., vol. 50, no. 1, pp. 46–62, Jan. 2006.
- [8] Y. Chabchoub, C. Fricker, F. Guillemin, and P. Robert, "On the statistical characterization of flows in internet traffic with application to sampling," Comput. Commun., vol. 33, no. 1, pp. 103–112, Jan. 2010.
- [9] Y. Shao, B. Yang, J. Jang, Y. Xue, and J. Li, "Emilie: Enhance the power of traffic identification," in Proc. ICNC, Feb. 2014, pp. 31–35.
- [10] S. Hegde, S. G. Koolagudi, and S. Bhattacharya, "Scalable and fair forwarding of elephant and mice traffic in software defined networks," Comput. Netw., vol. 92, pp. 330–340, Dec. 2015.
- [11] C. Wang, G. Zhang, H. Chen, and H. Xu, "An ACO-based elephant and mice flow scheduling system in SDN," in Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA), Mar. 2017, pp. 859–863.
- [12] M. Al-Saadi, A. Khan, V. Kelefouras, D. J. Walker, and B. Al-Saadi, "Unsupervised machine learning-based elephant and mice flow identification," in Intelligent Computing. Cham, Switzerland: Springer, 2021, pp. 357–370.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



SJIF Scientific Journal Impact Factor



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details